View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

Current Biology 22, 1456–1457, August 7, 2012 ©2012 Elsevier Ltd All rights reserved

Erratum

Multigene Phylogeny of the Green Lineage Reveals the Origin and Diversification of Land Plants

Cédric Finet,* Ruth E. Timme, Charles F. Delwiche, and Ferdinand Marlétaz

(Current Biology *20*, 2217–2222; December 21, 2010)

Following the publication of our multigene phylogeny of the green lineage, some studies have proposed a distinct branching pattern among algal relatives of land plants [1, 2]. Most notably, these differences relate to the position of the taxon *Coleochaete* that we identified as a land plant sister group and the failure to recover the monophyly of Coleochaetales in our study.

Recent exchanges of views with Hervé Philippe and colleagues have prompted us to investigate the possible cause of these inconsistencies. In particular, we searched for contaminations that had been identified by Philippe and colleagues ([3], this issue of *Current Biology*). We identified cross-contaminations in two libraries sequenced for our study, corresponding to the species *Chaetosphaeridium globosum* and to a lesser extent *Nitella hyalina*. The contamination likely took place during library preparation or sequencing, because the samples were obtained from pure in-house algal cultures, and there is a correlation between time of sample processing at the sequence facility and levels of cross-contamination. It is also noteworthy that the most highly contaminated data set, with as much as 5% exogenous sequence, was from *C. globosum*, which was the one data set that was amplified prior to the sequencing reactions. *C. globosum* and *N. hyalina* had what appear to be erroneous sequences for 28 out of 77 genes, representing 47 out of 5,929 sequences in the data matrix. Unfortunately, the phylogenetic proximity of these taxa and the close similarity of these sequences caused the contaminant sequences to escape our validation protocol.

To confirm the validity of our phylogenetic results, we set about evaluating the impact of the erroneous sequences on the reconstructed topology. We reconstructed independent trees with all significant hits (e value 1e−10) for the three concerned libraries and selected bona fide sequences among contaminants using a parsimony rule. We also included *Chara* data made recently available at the NCBI Sequence Read Archive [1]. We then reassembled corrected ribosomal protein data sets and performed subsequent phylogenetic inference using maximum-likelihood and Bayesian inference. In agreement with the limited extent of the contamination, the tree topology was largely unaffected; however, there was one very interesting rearrangement among the lineages most closely related to embryophytes. In the corrected analyses, the sister taxon to embryophytes is a monophyletic group comprising both Coleochaetales and Zygnematales, although a single branch swap would yield either

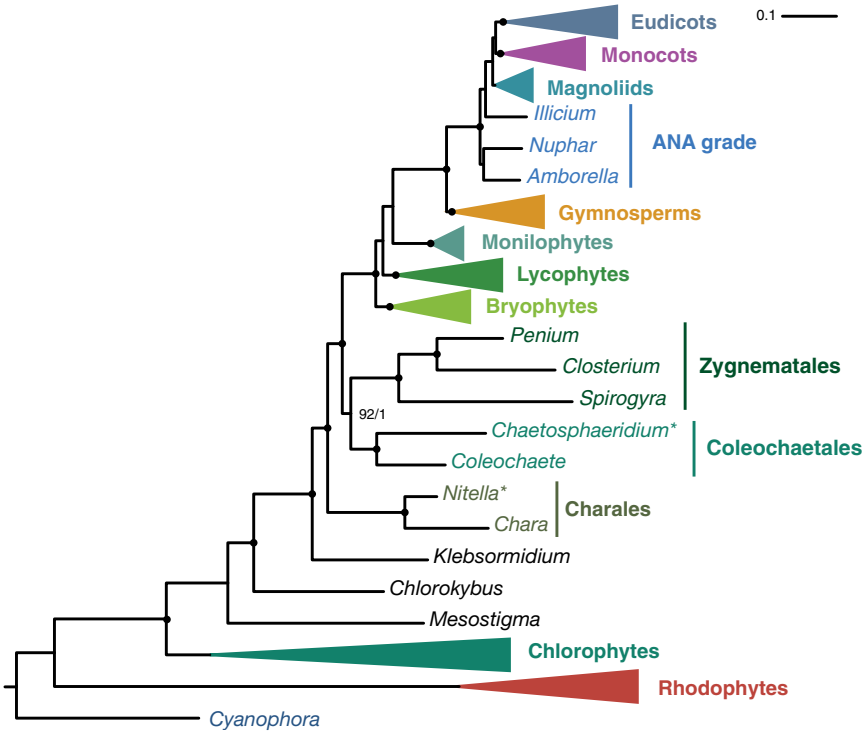

Figure 1. Phylogram of the Corrected 77-Taxon Analysis

RAxML maximum-likelihood analyses and PhyloBayes Bayesian analyses were conducted under the PROTMIXWAG model and the CAT model, respectively. Taxa marked with an asterisk (*) are those for which major contamination was observed and corrected. Support values obtained after 100 bootstrap replicates (BP) and Bayesian posterior probabilities (PP) are shown for selected branches. A bullet (●) indicates support values of BP = 100 and PP = 1.

the same topology seen in our original study (except for the placement of *Chaetosphaeridium*) or else the topology reported previously [1, 2] (Figure 1). Perhaps the most striking rearrangement is the emergence of a monophyletic Coleochaetales (i.e., *Coleochaete* and *Chaetosphaeridium* appear as sister taxa). We hypothesize that the possible introduction of *Penium* sequences into the *Chaetosphaeridium* cDNA library diminished the phylogenetic support for grouping *Coleochaete* and *Chaetosphaeridium* together.

Importantly, regarding character orientation, it is noteworthy that the revised position of *Coleochaete* is consistent with the major conclusion of our paper that relatively simple algal forms may be the closest relatives of land plants, and that the whorled branches observed in Charales likely evolved convergently, a conclusion that is also supported by previous analyses [1, 2]. Contamination has been a long-standing problem in molecular phylogenies and phylogenomics, especially when exploring new areas of biodiversity, and is particularly problematic in very large data sets that are difficult or impossible to fully curate manually. This case clearly pleads for further development of computational methods to detect contaminants in the growing amounts of sequence data generated for new taxa.

References

1. Wodniok, S., Brinkmann, H., Glöckner, G., Heidel, A.J., Philippe, H., Melkonian, M., and Becker, B. (2011). Origin of land plants: do conjugating green algae hold the key? BMC Evol. Biol. *11*, 104.
2. Timme, R.E., Bachvaroff, T.R., and Delwiche, C.F. (2012). Broad phylogenomic sampling and the sister lineage of land plants. PLoS ONE *7*, e29696.
3. Laurin-Lemay, S., Brinkmann, H., and Philippe, H. (2012). Origin of land plants revisited in the light of sequence contamination and missing data. Curr. Biol. *22*, this issue, R593–R594.

*Correspondence: cedric.finet@ens-lyon.org

http://dx.doi.org/10.1016/j.cub.2012.07.021